

# Sorting duplicated loci disentangles complexities of polyploid genomes masked by genotyping by sequencing

**Limborg, Morten; Seeb, Lisa W.; Seeb, J. E.**

*Published in:*
Molecular Ecology

*Link to article, DOI:*
10.1111/mec.13601

*Publication date:*
2016

*Document Version*
Publisher's PDF, also known as Version of record

MOLECULAR ECOLOGY

Molecular Ecology (2016) **25**, 2117–2129

NEWS AND VIEWS

OPINION

# Sorting duplicated loci disentangles complexities of polyploid genomes masked by genotyping by sequencing

MORTEN T. LIMBORG,*†§ LISA W. SEEB* and JAMES E. SEEB*

*School of Aquatic and Fishery Sciences, University of Washington, 1122 NE Boat Street, Box 355020, Seattle, WA 98195, USA; †National Institute of Aquatic Resources, Technical University of Denmark, Vejlsøvej 39, 8600 Silkeborg, Denmark*

**Many plants and animals of polyploid origin are currently enjoying a genomics explosion enabled by modern sequencing and genotyping technologies. However, routine filtering of duplicated loci in most studies using genotyping by sequencing introduces an unacceptable, but often overlooked, bias when detecting selection. Retained duplicates from ancient whole-genome duplications (WGDs) may be found throughout genomes, whereas retained duplicates from recent WGDs are concentrated at distal ends of some chromosome arms. Additionally, segmental duplicates can be found at distal ends or nearly anywhere in a genome. Evidence shows that these duplications facilitate adaptation through one of two pathways: neo-functionalization or increased gene expression. Filtering duplicates removes distal ends of some chromosomes, and distal ends are especially known to harbour adaptively important genes. Thus, filtering of duplicated loci impoverishes the interpretation of genomic data as signals from contiguous duplicated genes are ignored. We review existing strategies to genotype and map duplicated loci; we focus in detail on an overlooked strategy of using gynogenetic haploids (1N) as a part of new genotyping by sequencing studies. We provide guidelines on how to use this haploid strategy for studies on polyploid-origin vertebrates including how it can be used to screen duplicated loci in natural populations. We conclude by discussing areas of research that will benefit from better inclusion of polyploid loci; we particularly stress the sometimes overlooked fact that basing genomic studies on dense maps provides value added in the form of locating and annotating outlier loci or colocating outliers into islands of divergence.**

*Keywords*: genotyping by sequencing, haploid, homeologous recombination, isoloci, map-based genomics, polyploid, whole-genome duplication

*Received 9 July 2015; revision received 19 February 2016; accepted 24 February 2016*

## Background

Whole-genome duplication (WGD) is a fundamental evolutionary process known to accelerate adaptation (Ohno 1970b; Selmecki *et al.* 2015). Many polyploid-origin plants and animals, especially some crops, amphibians and fish, are currently enjoying a genomics revolution fueled by modern sequencing technology (Narum *et al.* 2013; Dolezel *et al.* 2014; Ellegren 2014; Shaffer *et al.* 2015). However, potentially large numbers of **paralogous**[1] loci in polyploid-origin genomes challenge genomic analyses; genome assemblies are particularly challenging due to equivocal sequence alignment. Polyploid-origin genomes undergoing **rediploidization** are affected by **homeologous** recombination. This results in **residual tetrasomic inheritance** that retards divergence of duplicated loci at the distal ends of some homeologous chromosomes (Fig. 1a; reviewed in Gaeta & Pires 2010; Allendorf *et al.* 2015; May & Delany 2015). As a result, the extent of rediploidization of duplicated loci following a WGD can vary across the genome, where some loci, **diverged duplicates**, are fully diploidized, while other duplicates, referred to as **isoloci** (Allendorf & Thorgaard 1984), actively share alleles by retained tetrasomic inheritance (Table 1).

Tetrasomic pairing and homeologous crossovers are reported in a range of autopolyploid taxa such as plants, yeast, amphibians and fish. For example, Stift *et al.* (2008) showed that the rate of residual tetrasomy differs among species and chromosomes in plants; the authors inferred homeologous recombinations from observations of inheritance ratios intermediate to disomic and tetrasomic in hybrids of the yellow cress plant (*Rorippa* spp). Similar inheritance patterns in the autotetraploid grey tree frog (*Hyla versicolor*) led Danzmann & Bogart (1982) to conclude that homeologous chromosomes occasionally pair during meiosis. Preferential pairing of homeologous chromosomes in hybrid genomes was postulated to be the cause of the unusual pattern of joint segregation of unlinked loci in autopolyploid-origin salmonids (pseudolinkage; Wright *et al.* 1983). All of these studies used experimental breeding

[1]Terms in boldface are explained in the glossary below.

Correspondence: James E. Seeb, Fax: 1 206 685 7471; E-mail: jseeb@uw.edu

§Present address: Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Copenhagen K, Denmark

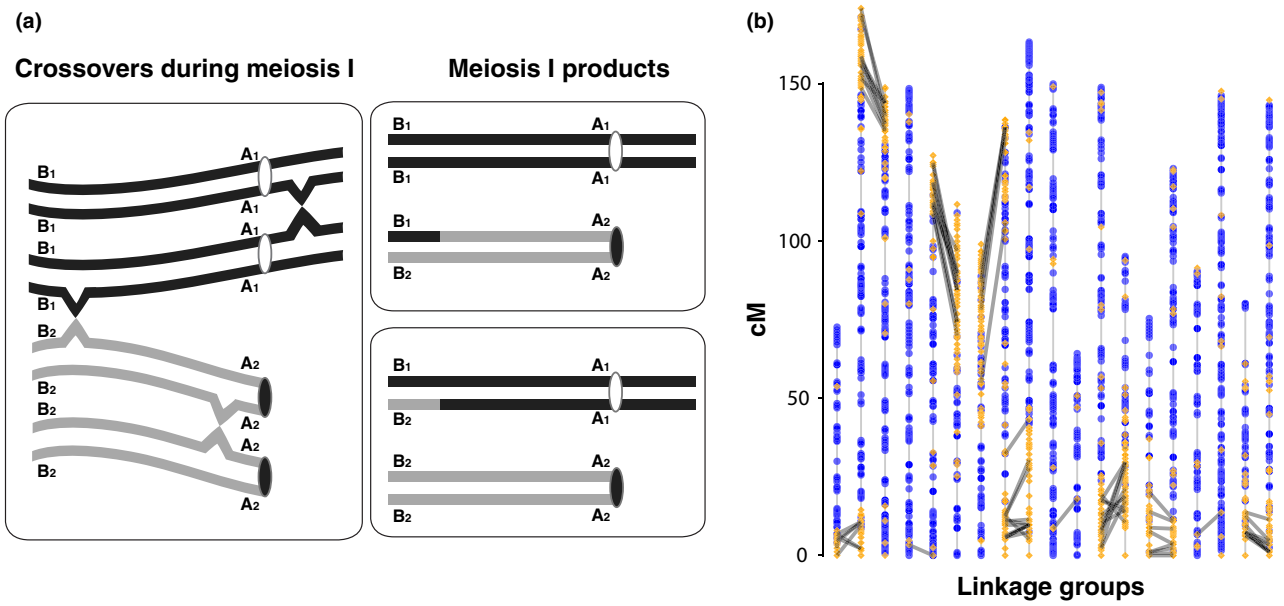


**Fig. 1** Model of residual tetrasomic inheritance and the resulting genomic distribution of duplicated loci. (a) Model of partial tetrasomic inheritance resulting from homeologous crossovers. *A*1 and *A*2 represent diverged duplicates that are located near their centromeres. *B*1 and *B*2 represent isoloci that share alleles. Crossovers that sometimes may occur between homeologs will produce intermediate disomic–tetrasomic segregation ratios for more distal loci (*B*1 and *B*2), while loci near the centromere (*A*1 and *A*2) will show disomic inheritance. Homeologs will pair with each other only in distal regions that retain sequence similarity. (b) A subset of 19 linkage groups from the autotetraploid-origin chum salmon (*O. keta*). Haploid-assisted mapping was used to map nonduplicated (blue circles) and duplicated (yellow diamonds) loci. Haploid-assisted mapping allowed the identification of homeologous chromosome pairs through the inclusion of both loci at some duplicate pairs (loci linked by lines between linkage groups). Areas dominated by duplicated loci are assumed to represent distal chromosomal regions with residual tetrasomic inheritance. Figure 1b redrawn from Waples *et al.* (2016) with permission.

to disentangle the complex inheritance patterns observed in these polyploid-origin genomes.

Here, we clarify current challenges for conducting population genomics in polyploid-origin species. We briefly review the current strategies and limitations of widely used genotyping by sequencing (GBS) techniques and highlight shortcomings of studying unmapped polyploid genomes where duplicated loci are filtered. Next, we introduce salmonids as a model for studying polyploid-origin vertebrates by using existing protocols to reduce genome complexity. We then provide guidelines on how haploid-assisted mapping can help to identify, map and subsequently genotype duplicated loci including isoloci. We end by presenting general areas where future studies that include duplicated loci are expected to enhance evolutionary insights and improve the conservation of polyploid-origin species.

## Genotyping by sequencing fails to resolve duplicated loci

Novel genomic-based methods have rapidly accelerated SNP discovery and acquisition of genome-wide data sets (Andrews *et al.* 2016). GBS is a relatively new and especially popular technique that does not require existing genomic resources, making it particularly suitable for nonmodel organisms. GBS targets a reduced representation of a genome that is then sequenced with next-generation sequencing to obtain genotypes at thousands of loci (Davey *et al.* 2011). Use of GBS enabled genome mapping and QTL studies in a range of nonmodel taxa (e.g. Barchi *et al.* 2012; Houston *et al.* 2012; Narum *et al.* 2013); dense meiotic maps have provided a powerful framework for studying the genomic architecture of divergence in natural populations (Baxter *et al.* 2011; Chutimanitsakun *et al.* 2011) including polyploid-origin salmonids (Gagnaire *et al.* 2012; Hecht *et al.* 2013).

A significant problem is that many genotyping techniques, including GBS (Davey *et al.* 2011), are unable to properly genotype duplicated loci (Krasileva *et al.* 2013; Dufresne *et al.* 2014; Dufresne 2016). Most researchers using GBS approaches actively filter such duplicates in mapping and population studies in order to acquire data sets that meet standard assumptions such as Mendelian inheritance and Hardy–Weinberg equilibrium (e.g. Seeb *et al.* 2011; Gagnaire *et al.* 2012; many others). However, this filtering leads to the exclusion of **segmental duplicates** as well as sometimes large portions of the distal ends of a subset of chromosomes in polyploid-origin species (Kodama *et al.* 2014; McKinney *et al.* 2015; Waples *et al.* 2016), a fact that is rarely acknowledged in studies interpreting genome-wide results of both plants (reviewed

**Table 1** Different states of duplicated loci during rediploidization after whole-genome duplication. Within a rediploidizing genome, some loci can be fully diploidized, while other loci maintain partial tetrasomic inheritance depending on their location in the genome (Fig. 1)

| Locus type | Duplication status | Inheritance | Definition* |
|---|---|---|---|
| Isolocus | Duplicated | Residual Tetrasomic | A pair of genetically distinct loci that still share identical alleles (from ongoing, or recent, tetrasomic inheritance); alleles cannot be unambiguously assigned to a single locus |
| Diverged duplicates | Duplicated | Disomic | Different alleles have become 'fixed', or nearly so, at the two duplicates |
| Single locus | Nonduplicated | Disomic | Only one gene copy exists in the genome resulting from either loss or functional divergence of the other duplicate |

*Based on nomenclature and definitions given in Allendorf & Thorgaard (1984).

in Mastretta-Yanes *et al.* 2014) and animals (e.g. Gagnaire *et al.* 2012; Limborg *et al.* 2014; Leaché *et al.* 2015; many others).

Further, alleles at SNP loci located in duplicated regions resulting from homeologous crossovers cannot easily be assigned to their specific chromosomal origin when scored using GBS (reviewed in Dufresne *et al.* 2014; Mastretta-Yanes *et al.* 2014; Logan-Young *et al.* 2015). Waples *et al.* (2016) inferred that residual tetrasomy (Fig. 1a; Allendorf *et al.* 2015; May & Delany 2015) maintains large blocks of undiverged duplicated loci, isoloci, at the distal ends of some chromatids in chum salmon (*Oncorhynchus keta*; Fig. 1b). Therefore, routine exclusion of such duplicates from mapping and population studies biases results in a way that could misinform some interpretations. In humans, for example, subtelomeric regions are known to be complex, dynamic and variable regions that harbour adaptively important gene families (Mefford *et al.* 2001; Mefford & Trask 2002). Numerous telomere-associated genes in mammals have also been shown to be adaptively important (Morgan *et al.* 2013). Clearly, studies that filter duplicated loci may be missing genomic regions of adaptive importance.

Another important and related challenge is that GBS data are blind to allele dosage except in the atypical case of ultra-deep sequencing (Lighten *et al.* 2014). Allozyme markers, which were originally used to describe some duplicated loci, revealed allelic dosages by amplifying equally expressed products of four allele copies, facilitating the detection of duplicated loci with the naked eye (Box 1; Allendorf & Thorgaard 1984). Similarly, fluorescence data specific to some SNP calling platforms enabled the scoring of dosage in duplicated SNPs (Box 1; Gidskehaug *et al.* 2011; Voorrips *et al.* 2011). In contrast, inference of dosage remains equivocal in typical GBS data sets where diploid heterozygotes [AB] cannot easily be distinguished from heterozygotes at one or both pairs of duplicated loci that amplify four allele copies [AA/AB or AB/BB], double heterozygotes [AB/AB] or alternate homozygotes [AA/BB] (see also discussions in Dufresne *et al.* 2014; Waples *et al.* 2016). When duplicated loci are either filtered or treated as diploid, downstream analyses will be biased potentially leading to erroneous interpretations (Allendorf *et al.* 2015). Hence, initial discrimination between duplicated and nonduplicated loci based on GBS data alone must rely on arbitrary thresholds of sequence coverage, sequence similarity and statistical fit to disomic models (e.g. Hardy–Weinberg; see Box 1 for further discussion).

## Salmonids as models for studying polyploid vertebrates

Protocols for performing experimental breeding in vertebrates are generally much more cumbersome than protocols for plants. Duplicated loci are more easily resolved in plants because often-simple husbandry easily enables the implementation of inbred lines, backcrosses and **doubled haploids** (e.g. Ogugo *et al.* 2015; Siger *et al.* 2015); as a result, economically important crops sometime serve as models for studying genome evolution (see Wallace *et al.* 2014). Chromosome set manipulations such as the production of doubled haploids aided mapping of polyploid plant genomes (e.g. Poland *et al.* 2012; Chapman *et al.* 2015; Fan *et al.* 2015), and tissues in plants are often readily accessible from haploid life stages (e.g. Fins & Seeb 1986; Richier *et al.* 2009), enabling unambiguous genotyping and map-based genomic study (Eckert *et al.* 2010; Sork *et al.* 2013). Unfortunately, reproductive biology and potentially complicated husbandry of many vertebrates greatly limit the execution of similar breeding and genotyping protocols. Here, we focus on how to advance studies in vertebrates.

Salmonids provide an ideal system for addressing challenges with analyses of duplicated genomes in vertebrate species. A common ancestor to salmonids went through a WGD (Ohno 1970a) dated to 88-103 MYA (Macqueen & Johnston 2014), and the salmonid genome is still undergoing rediploidization. Numerous inheritance studies using allozyme loci during the 1970s and 1980s demonstrated tetrasomic inheritance patterns in males for pairs of isoloci that remain undiverged and thus share alleles (Wright *et al.* 1983; Allendorf & Thorgaard 1984; Allendorf & Danzmann 1997). Importantly, isoloci were shown to cluster at the distal ends of chromosomes both from allozyme (Thorgaard *et al.* 1983; Allendorf *et al.* 1986; Seeb & Seeb 1986) and subsequently from microsatellite data (Sakamoto *et al.* 2000; Guyomard *et al.* 2012). These often large syntenic blocks of duplicated genes are generally found on a subset

**Box 1.** Comparing allozymes, GBS and SNP arrays: The issue with dosage

The ability to infer allele dosage is required for de novo detection and screening of duplicated loci. An inherent hurdle for understanding duplicated genomes using genotyping by sequencing (GBS) is that, in contrast to allozymes, allele dosage (i.e. allele copy number) is not directly conveyed (Fig. A).

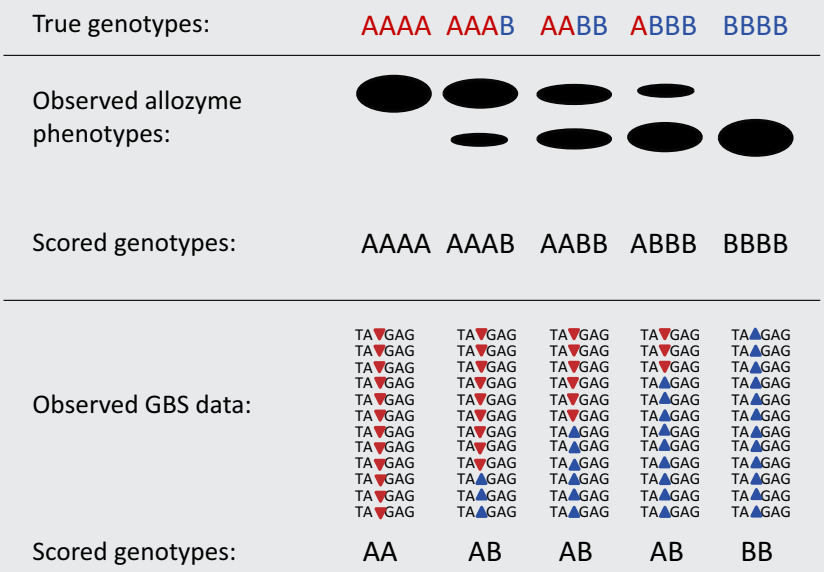


**Fig. A.** Scoring allele copies at a duplicated locus using allozyme or GBS markers. Downward and upward pointing triangles represent alternate nucleotides at a SNP locus and link to the A (red) and B (blue) allele of the true genotypes listed above. Allozyme markers signal allele dosage based on the ratios of observed gel bands. In contrast, inference of allele dosage remains a major challenge with typical GBS data. This inevitably results in misclassification of some duplicated loci that will be scored as disomic heterozygotes because allelic ratios of 1:3 cannot be distinguished from ratios of 2:2 with routine sequence coverage.

of eight conserved pairs of chromosome arms in more recently published high-density maps of several salmonid species (Fig. 1b; Gidskehaug *et al.* 2011; Brieuc *et al.* 2014; McKinney *et al.* 2015; Waples *et al.* 2016).

Genomic resources for salmonids are exceedingly common further facilitating genomic studies (Thorgaard *et al.* 2002; Davidson *et al.* 2010). Their life history (especially anadromy and fine-scale homing) has made them ideally suited for using genetic data to improve population management (e.g. Dann *et al.* 2013), improve fish culture (e.g. Allen *et al.* 2014) and advance our understanding of local adaptation in the wild (reviewed in Fraser *et al.* 2011). Further, fecundity of salmonids readily enables studies of transmission genetics and gene mapping: single-pair matings yield 100s to 1000s of embryos, permitting the enumeration of large numbers of meioses. Finally, details of **ploidy manipulation** are well worked out (Komen & Thorgaard 2007), and the use of **haploid families** has been shown to greatly enhance the characterization of duplicated loci (e.g. Spruell *et al.* 1999; Lien *et al.* 2011; Waples *et al.* 2016).

## Linkage mapping based upon chromosome set manipulations offers a solution

The use of **chromosome set manipulations** can greatly simplify the identification and mapping of duplicated loci. A variety of chromosome set manipulations produce haploids, triploids or inbred diploids by the use of either or both of two steps: (i) irradiation to enucleate spermatozoa or ova and (ii) blocking a cell division with a thermal or pressure shock to disrupt the spindle apparatus (Komen & Thorgaard 2007; Abu-Daya *et al.* 2012; Box 2). These manipulations can be implemented easily in many species with external fertilization including amphibians (e.g. Selman 1958; Kondo & Kashiwagi 2004), fish (e.g. Kroeger *et al.* 2014; Xu *et al.* 2014; Molina-Luzon *et al.* 2015) and molluscs (e.g. Li *et al.* 2004; Bottger *et al.* 2011).

Meiotic maps generated from families of **gynogenetic haploid** individuals provide several very important advantages, particularly in polyploid-origin species (Spruell *et al.* 1999; Waples *et al.* 2016). Haploids only contain a single gene copy at diploid loci, so the presence of two or more different alleles at a locus signals a duplication where alleles from both duplicates have been assembled into a single locus due to retained sequence similarity. Previous GBS-based studies have used this asset to either filter duplicated loci from QTL and population genetic studies (Everett & Seeb 2014; Limborg *et al.* 2014) or map duplicated loci (Brieuc *et al.* 2014; Waples *et al.* 2016). Another benefit of haploids is that they contain only half of the DNA of diploids, reducing sequencing costs.

**Box 1** *continued*

Gidskehaug *et al.* (2011) presented an approach that allows genotyping thousands of markers while considering allele dosage at duplicated loci (Fig. B). The relative strengths of fluorescent signals from two allele-specific bead arrays, using Illumina's Infinium® technology, allowed inference of dosage at duplicated SNP loci. Paralogous loci were then classified as multisite variants (MSV); when one paralog was homozygous and the other heterozygous, three distinct genotype clusters were distinguished (MSV-3). Five genotypes were distinguished when both paralogs were heterozygous (MSV-5). This technique was used to presents one of the first efforts to map duplicated loci on a linkage map of Atlantic salmon (*Salmo salar*) illustrating subtelomeric clustering of duplicated loci (Lien *et al.* 2011).

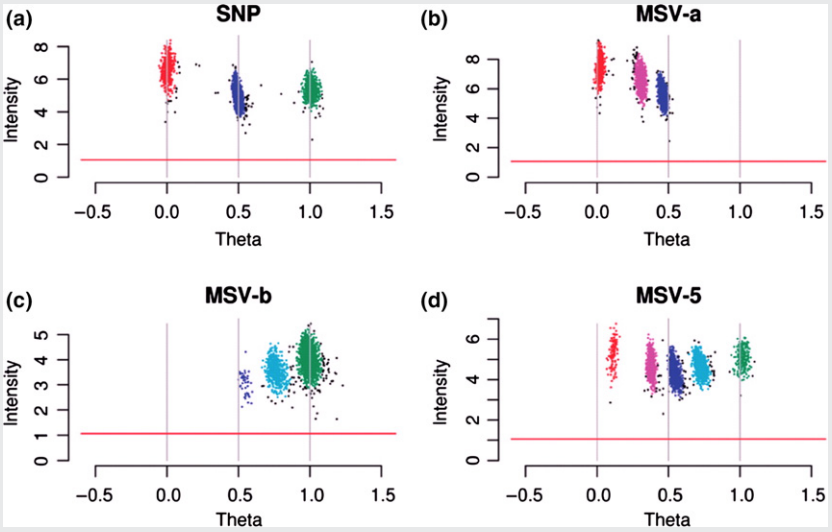


**Fig. B.** Examples of fluorescence-based genotype clustering for different marker categories. (a) A biallelic SNP (strength of fluorescence is measured by the theta parameter on the *x*-axis). (b–c) The MSV-3 type of a duplicated marker where only one of the duplicates is polymorphic; the three clusters conform to allelic ratios of: 4:0 (AAAA), 3:1 (AAAB) and 2:2 (AABB). (d) An MSV-5 duplicated marker where both duplicates are polymorphic results in five clusters corresponding to all combinations of the two alleles: 4:0 (AAAA), 3:1 (AAAB), 2:2 (AABB), 1:3 (ABBB) and 0:4 (BBBB). The figure is reproduced from Gidskehaug *et al.* (2011) with permission.

Currently, fluorescent signals to genotype duplicated SNPs are not commonly used, although this should be possible (Gidskehaug *et al.* 2011; Johnston *et al.* 2013). A more immediate drawback of this approach is that the generation of species-specific SNP arrays for thousands of loci is an expensive and time-consuming task that makes this approach exceedingly challenging for many nonmodel species compared to the immediate genotyping available through GBS. Unfortunately, many existing SNP arrays exclude duplicated loci. For example, the major effort that produced the 57K fluorescent-based array for rainbow trout was made after most duplicated loci had been actively filtered (Palti *et al.* 2015).

**Gynogenetic diploids** have also been shown to be extremely useful in genome mapping. All gynogenetic diploids possess recombination information because they are **half-tetrads** that contain both sister chromatids from meiosis II (Thorgaard *et al.* 1983). As a result, observed recombination frequencies provide direct estimates of gene–centromere distances (Thorgaard *et al.* 1983; Seeb & Seeb 1986; Lindner *et al.* 2000). Gynogenetic diploids have more recently been used to locate centromeres on medium-density (Sakamoto *et al.* 2000; Gharbi *et al.* 2006) and high-density (Brieuc *et al.* 2014; Limborg *et al.* 2015) linkage maps.

Although powerful for mapping, most of the attention given to doubled haploids in fish has its origin in the utility of isogenic lines for aquaculture and breeding studies (cf., Thorgaard 1992; Grimholt *et al.* 2009). Doubled haploids are technically difficult to produce compared to haploids, and low survival rates impede the generation of large mapping families (cf., Parsons & Thorgaard 1985). However, doubled haploids have been instrumental in the projects to assemble whole-genome sequences of Atlantic salmon (Davidson *et al.* 2010; Davidson 2013) and rainbow trout (*O. mykiss*; Berthelot *et al.* 2014); these projects relied on just a single doubled haploid individual for sequencing.

## Approaches for genotyping duplicated loci in natural populations

Using haploids to identify and map duplicated loci is a practical first step towards genotyping duplicated loci in

**Box 2.** Chromosome set manipulations

We briefly review different chromosome set manipulations in fish to clarify the advantages and difficulties of each and provide a clearer understanding of the simplicity of mapping polyploid-origin species using haploid gynogenesis. These methods work well in any organism with external fertilization, such as most fish and amphibian species.
Gynogenetic offspring can be created by the inactivation of the DNA in the sperm prior to fertilization (Thorgaard *et al.* 1983). Sperm DNA is readily denatured by exposure to ultraviolet (UV) light; activation of ova using irradiated spermatozoa produces gynogenetic haploid zygotes (Fig. C). Haploid embryos survive for many weeks, providing DNA for family genotyping, but haploids die prior to hatch. The second technique, formation of gynogenetic diploids, is possible because meiosis in fish ova is temporarily arrested in anaphase II, and the application of a shock to gynogenetic haploids shortly after fertilization disrupts cell division in meiosis II. That treatment incorporates the second polar body into the nucleus, forming gynogenetic diploids (Fig. C). Survival of gynogenetic diploids varies, but typically is adequate for most experimental studies (Allendorf *et al.* 1986). Alternatively, the application of a shock to gynogenetic haploids during first mitosis disrupts that cell division and creates gynogenetic doubled haploids (Fig. C). Survival of doubled haploid salmonids is exceedingly low (Komen & Thorgaard 2007).

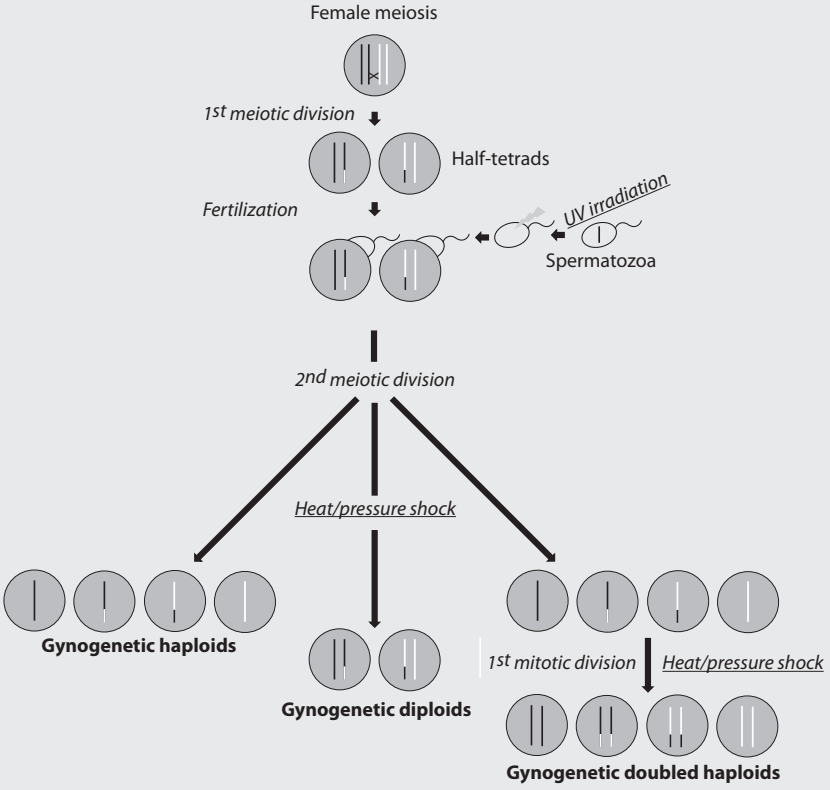


**Fig. C.** Chromosome set manipulations for generating gynogenetic haploids, gynogenetic diploids and gynogenetic doubled haploids. UV irradiation of spermatozoa prior to fertilization ensures that only maternal DNA is incorporated into developing embryos. Haploids are generated with no further treatment. A shock to disrupt meiosis II interrupts extrusion of the second polar body and generates gynogenetic diploids, and a shock to disrupt first mitosis in haploids generates fully homozygous doubled haploids.

Androgenetic haploids can be generated in a similar, but much more technologically challenging treatment, by gamma irradiation of ova followed by fertilization with normal spermatozoa. Disruption of the first mitosis will generate androgenetic doubled haploids (Fig. D). These diploid androgens have a 1:1 ratio of males to females, but again, survival is low (Parsons & Thorgaard 1985). Doubled haploids generated from the heterogametic parent will create both male and female offspring; in salmonids for example, males are the heterogametic sex, so androgenesis was routinely used if multigenerational doubled haploid lines were needed because both male and female offspring were produced (Scheerer *et al.* 1991; Thorgaard 1992).

**Box 2** *continued*

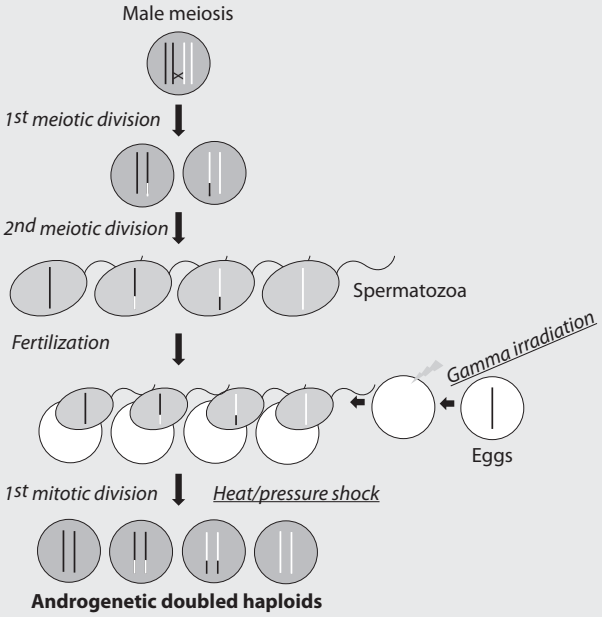


**Fig. D.** Chromosome set manipulations for generating androgenetic doubled haploids. Gamma irradiation is used to enucleate female gametes to ensure that only paternal DNA is incorporated into developing embryos. Shock to disrupt first mitosis generates doubled haploids.

natural populations of polyploid organisms. A promising example is given in Waples *et al.* (2016) who used a maximum-likelihood model to characterize the segregation of duplicated loci in haploid families generated through chromosome set manipulation. These types of efforts will help standardize methods to identify and map duplicated loci across laboratories and taxa.

Improved sequencing technologies now provide longer read lengths (e.g. Mosher *et al.* 2014) that will improve the ability to distinguish diverged duplicates and facilitate direct genotyping of these loci (e.g. Clevenger & Ozias-Akins 2015). Diverged duplicates residing in chromosomal regions that exhibit only disomic inheritance have accumulated novel mutations and will be easier to differentiate with longer reads. Such improvements will be particularly useful for species where mapping is difficult (e.g. many birds and mammals with low fecundities). Unfortunately, isoloci with residual tetrasomic inheritance will remain unresolved because the sharing of alleles prevents unambiguous mapping, regardless of read length, and isoloci will have to be considered as single tetraploid loci.

A promising alternative is to use a dominant genotyping model that can provide a pathway to more immediate insights. A dominant genotype can be inferred by scoring the presence or absence of alleles at duplicated loci (e.g. Tomiuk *et al.* 2009; Dufresne *et al.* 2014). Dominance approaches have been successfully applied with duplicated microsatellite markers for parentage analyses in polyploid sturgeon (Wang & Scribner 2014) and appear promising for obtaining conservative estimates of population divergence at isoloci. One drawback of dominant genotyping is its reduced statistical power because it does not control for allelic dosage uncertainty at heterozygote genotypes (Box 1). A potential solution is to consider the number of sequence reads at individual loci to estimate allele frequencies in a population sample that can then be used for analyses of population differentiation and genome scans for selection (see Blischak *et al.* 2015).

Many nonmodel species are expected to have a genome reference sequence available in the foreseeable future. This will facilitate genotyping through whole-genome resequencing allowing denser genome sampling compared to GBS. Nevertheless, assembling a genomic reference sequence in duplicated regions exhibiting residual tetrasomic inheritance is not trivial (see Rondeau *et al.* 2014). Therefore, analyses of isoloci will continue to rely on chromosome set manipulations for distinguishing diverged diploid regions from homeologous regions with residual tetrasomic inheritance. It will be possible to more effectively study all duplicated regions with whole-genome resequencing once a genomic reference sequence correctly accounts for retained homeology. The ability to perform whole-genome resequencing will also mitigate some limitations presented by GBS including issues relating to missing data (Andrews *et al.* 2016).

Finally, it is important to note that most of these methods require an *a priori* assumption about duplication status, or ploidy, of each locus (cf., Table 1). Haploid individuals provide a pathway to obtain this information for many nonmodel species. These genotyping approaches may then

enable the scoring of dosage at duplicated loci in populations, which is useful, but alone they will not provide the genotypes of individual loci in pairs of isoloci. This necessitates the consideration of isoloci as single tetraploid loci when interpreting these gene duplicates in a population genetic framework.

## Future perspectives

We believe that increased focus on using available techniques such as haploids to resolve duplicated genes will increase the inclusion of these loci in population genomics studies in the foreseeable future; it is therefore not too early to ignite excitement about potential studies of species with polyploid genomes. Below, we discuss specific questions that can help elucidate the evolutionary role of genome, and local gene duplications, as well as improve management and conservation of polyploid-origin species in general. Obviously, this list is far from exhaustive, but we hope our examples will spark further ideas across all polyploid taxa.

1 *Map-based population genomics*. When a high-quality reference genome is not available, which is still the case for most nonmodel species, linkage maps provide a basis to improve the interpretation of population genomic data. Studies that complement genome scans with linkage maps demonstrate increased power to detect biologically meaningful signals of adaptation in both animals (Hemmer-Hansen *et al.* 2013; Santure *et al.* 2013) and plants (Eckert & Dyer 2012; Stölting *et al.* 2013; see also discussion in Limborg *et al.* 2014). Here, we highlight the need for proper genotyping of duplicated loci as well as provide an improved pathway to map-based genomics in polyploid-origin vertebrates.

2 *Evolutionary role of gene duplications.* Genome duplications are thought to catalyse evolutionary processes and play a key role in speciation events (Ohno 1970b; Wolfe 2001; Comai 2005; Mable *et al.* 2011). In a recent study, Wu *et al.* (2015) showed that duplicates of genes involved with stress response in the green alga (*Chlamydomonas reinhardtii*) tended to have both copies retained more often than genes that are not involved with stress response. These results illustrate that gene duplications have played an important adaptive role in shaping green alga's resiliency to stress. The ability to interrogate extensive, and hitherto ignored, distal regions dominated by duplicated genes will expedite insights into the functional fates of gene duplicates (cf., Allendorf 1978).

3 *Adaptive potential of duplicated genes*. Isoloci are expected to have larger effective population sizes than disomic loci which will increase their effective response to selection (Charlesworth 2009). This leads to the expectation that genes with tetrasomic inheritance will show signatures of selection more often than disomically inherited genes. This hypothesis can now be tested, and insights will add to the ongoing debate on why some polyploids have been evolutionarily successful in contrast to diploid relatives (Allendorf & Thorgaard 1984; Comai 2005; Parisod *et al.* 2010; but see also Mayrose *et al.* 2011).

4 *Evolutionary success of polyploids.* The use of haploids can further add to the long-standing debate on the actual mechanisms explaining why many polyploids are successful (Ramsey & Schemske 1998). For instance, one hypothesis for the apparent success of many polyploid plants is their high origination rate allowing more opportunities for the polyploid variant to adapt and become successful (Otto & Whitton 2000). The use of haploids will allow future studies to consider a broader range of species, and this will undoubtedly help elucidate the general importance of origination rate and other mechanisms.

5 *Origin of duplicated genes.* Knowing the type of origin and genomic location of gene duplicates is essential for understanding the evolution of duplicated genes. Warren *et al.* (2014) detected paralog pairs from the Atlantic salmon transcriptome and considered map location to infer whether paralogs originated from a WGD (paralogs located on different chromosomes) or from a segmental gene duplication (paralogs located on the same chromosome). Whereas Warren *et al.* (2014) only considered diverged duplicates (see Table 1), techniques reviewed here promise to further improve such insights by allowing the inclusion of isoloci, assuring more complete coverage of the genome (Waples *et al.* 2016).

6 *Early stages of polyploidy in plants*. Many plants repeatedly create new autopolyploids where little to no intragenome differentiation occurs in recently formed populations (Parisod *et al.* 2010). The use of haploids with genome-wide linkage mapping has the potential to reveal segregation patterns across genomes in recently formed autopolyploids (Stift *et al.* 2008). This is a promising way to directly detect and distinguish loci with disomic or polysomic inheritance. Using this approach, it may be possible to better distinguish recently formed populations from more ancient polyploids and add to our understanding of the earliest stages of genomic and allelic diversification in newly formed polyploid lineages.

7 *Genomic architecture of adaptive divergence.* Subtelomeric regions are known to harbour genes of adaptive importance (Mefford & Trask 2002; Morgan *et al.* 2013). Interestingly, a recent study performing whole-genome resequencing in two flycatcher species (*Ficedula* spp.) revealed a heterogeneous genomic landscape with an overrepresentation of divergence peaks 'in the very end of chromosomes' (Ellegren *et al.* 2012). Hohenlohe *et al.* (2010) observed elevated nucleotide diversity at the ends of chromosomes in threespine stickleback (*Gasterosteus aculeatus*), although this diversity was not linked to population divergence. Genes important for host–pathogen interaction were also found to reside in subtelomeres (Kooij *et al.* 2005) or at the ends of chromosomes (Cornejo *et al.* 2015) in *Plasmodium* species. The degree to which these patterns are conserved across taxa, including polyploids, can only be tested with proper representation of the entire distal regions of all chromosomes.

8 *Inbreeding and outbreeding depression*. Resolving the effects of inbreeding and outbreeding depression is a classic theme in the study of genetics of natural populations. Inbreeding depression occurs when increased drift in small populations leads to either an increase in deleterious alleles or loss of variation at loci exhibiting heterozygote advantage. It can be speculated that gene duplicates that have retained the same function, for example all isoloci, will be more resilient to inbreeding depression from recessive mutations than will nonduplicated or functionally diverged gene duplicates (Allendorf & Thorgaard 1984; Comai 2005). Alternatively, outbreeding depression can occur from two different processes: reduced frequency of locally adapted alleles or from breakdown of coadapted genes. It remains unclear whether disomically inherited genes are differentially affected than isoloci exhibiting tetrasomic inheritance in autopolyploids (Parisod *et al.* 2010). Future insights may reveal more about the genomic mechanisms responsible for either susceptibility or resiliency to inbreeding and outbreeding depression in polyploid-origin species.

9 *Hybridization and invasive alleles.* Another outstanding question relates to invasive alleles in hybridized populations (Fitzpatrick *et al.* 2010). Invasive alleles have also been detected in populations of native westslope cutthroat trout (*O. clarkii*) hybridized with invasive rainbow trout (Hohenlohe *et al.* 2013). Considering that these hybrids experience reduced fitness (Muhlfeld *et al.* 2009), it will be interesting to assess whether invasive alleles are more likely to introgress in regions exhibiting tetrasomic inheritance (Chapman & Abbott 2010). Indeed, if an invasive allele conveys a fitness advantage by virtue of a dominant mutation, it is expected to be more often incorporated at a tetrasomic locus than at a disomic locus (Allendorf & Thorgaard 1984). The ability to map invasive alleles to both nonduplicated and duplicated genomic regions will help answer this question and increase our understanding of the genetic mechanisms affecting fitness in hybridized populations. Use of gene markers with known invasive alleles can be used for early detection of ongoing hybridization and guide conservation efforts towards populations at the highest risk of suffering from outbreeding depression.

## Conclusions

Herculean efforts were marshalled to sort duplicated regions in the genome assemblies of some polyploid organisms. For example, the entire genomes of preduplication ancestors were sequenced in order to sort the homeologs in wheat (*Triticum sp.*; Krasileva *et al.* 2013) and Atlantic salmon (Rondeau *et al.* 2014). Many tools have been applied to explore the polyploid genomes of plant species (Aversano *et al.* 2012); central to these and possibly the most powerful remains the genome map (Poland & Rife 2012; Mascher *et al.* 2013).

Yet application of GBS for genotyping and mapping polyploid animals and some polyploid plants has lagged compared to the rapid advances in diploid organisms (Bogart & Bi 2013). Single-cell sequencing is an interesting new technology potentially allowing direct sequencing of gametes and other haploid cells (Huang *et al.* 2015). However, for most nonmodel species, chromosome set manipulation combined with linkage mapping provides a technologically straightforward and robust approach and remains the most feasible method to interrogate gene duplicates that share alleles. For example, duplicated regions in polyploid amphibians remain unsorted (e.g. Robertson & Cornman 2014; Savage *et al.* 2014); haploid-assisted mapping could provide a substantial progress in a single generation.

We believe that future efforts interrogating duplicated genes, including haploid-assisted mapping and GBS, will advance our understanding of the evolutionary role of genome and gene duplications. Lessons from these studies will lead to an increased appreciation of the underlying genetic mechanisms of adaptation and translate into improved management of natural populations.

## Acknowledgements

We appreciate discussions during the 2014 American Genetic Association Workshop *Escape from Homeolog Hell* that sharpened our understanding of the intricacies of homeologous recombination. The spirited exchanges among Fred Allendorf, Roy Danzmann, Gary Thorgaard and Bernie May were especially helpful. Tyler Dann, Ryan Waples, Fred Allendorf and five anonymous reviewers provided useful comments on previous versions of this manuscript. MTL was supported by The Danish Council for Independent Research's career programme Sapere Aude Grant 12-126687 and a postdoctoral fellowship from the Carlsberg Foundation CF15-0721. JES and LWS were supported by grants from the Alaska Sustainable Salmon Fund, NOAA Saltonstall-Kennedy Grant Program and the State of Washington Sea Grant.

## Glossary

**Chromosome set (ploidy) manipulations:** A suite of breeding techniques that can be implemented to modulate the DNA content of resulting embryos to deviate from a classic diploid cross between a male and a female parent.

**Diverged duplicates (See Table 1):** Different alleles have become 'fixed', or nearly so, at the two duplicate genes.

**Doubled haploid:** Diploid individual formed by doubling the chromosomes of a haploid gamete. Doubled haploid progeny are fully homozygous.

**Gynogenetic diploid:** A diploid individual that contains both products from the first meiotic division from a single meiosis of the female parent.

**Gynogenetic haploid:** A haploid individual that only contains a single set of chromosomes inherited from the female parent.

**Half-tetrads:** A single progeny with two parental meiotic products recovered from the same meiosis.

**Haploid family:** A full-sib family where all progeny have only inherited haploid DNA from a single parent.

**Homeologous:** Pairs of chromosomes originating from the same ancestral chromosome before a whole-genome duplication. Homeology destabilizes meiosis; recombination can occur among different combinations of homeologs leading to improper disjunction. Divergence of homeologs during rediploidization re-introduces meiotic stability and starts from the centromere, moving distally. Residual tetrasomic inheritance will retain gene duplication in the distal ends of ancestral homeologs in otherwise fully rediploidized taxa.

**Isoloci (See Table 1):** A pair of duplicated loci that still share alleles from ongoing, or recent, tetrasomic inheritance. Alleles at isoloci cannot be unambiguously assigned to a single locus.

**Paralogous:** Homology between two DNA segments (paralogs) in the same genome originating from a duplication event. Includes both whole-genome and segmental duplications.

**Rediploidization:** The process whereby a duplicated genome reverts back to a diploid state. Rediploidization is hindered by recombination among homeologs and may occur at different rates throughout duplicated genomes.

**Residual tetrasomic inheritance:** Inheritance of two loci where observed segregation ratios are intermediate between those expected with disomic and tetrasomic inheritance. This results when occasional recombination between homeologs occurs.

**Segmental duplicates:** A common feature of many genomes resulting from a locally restricted DNA duplication not related to a whole-genome duplication event.

## References


Abu-Daya A, Khokha MK, Zimmerman LB (2012) The Hitchhiker's guide to *Xenopus* genetics. *Genesis*, **50**, 164–175.

Allen MS, Ferguson MM, Danzmann RG (2014) Molecular markers for variation in spawning date in a hatchery population of rainbow trout (*Oncorhynchus mykiss*). *Marine Biotechnology*, **16**, 289–298.

Allendorf FW (1978) Protein polymorphism and the rate of loss of duplicate gene expression. *Nature*, **272**, 76–78.

Allendorf FW, Danzmann RG (1997) Secondary tetrasomic segregation of MDH-B and preferential pairing of homeologues in rainbow trout. *Genetics*, **145**, 1083–1092.

Allendorf FW, Thorgaard GH (1984) Tetraploidy and the evolution of salmonid fishes. In: *Evolutionary Genetics of Fishes* (ed. Turner BJ), pp. 1–53. Plenum Press, New York.

Allendorf FW, Seeb JE, Knudsen KL, Thorgaard GH, Leary RF (1986) Gene-centromere mapping of 25 loci in rainbow-trout. *Journal of Heredity*, **77**, 307–312.

Allendorf FW, Bassham S, Cresko WA *et al.* (2015) Effects of crossovers between homeologs on inheritance and population genomics in polyploid-derived salmonid fishes. *Journal of Heredity*, **106**, 217–227.

Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA (2016) Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, **17**, 81–92.

Aversano R, Ercolano MR, Caruso I *et al.* (2012) Molecular tools for exploring polyploid genomes in plants. *International Journal of Molecular Sciences*, **13**, 10316–10335.

Barchi L, Lanteri S, Portis E *et al.* (2012) A RAD tag derived marker based eggplant linkage map and the location of QTLs determining anthocyanin pigmentation. *PLoS ONE*, **7**, e43740.

Baxter SW, Davey JW, Johnston JS *et al.* (2011) Linkage mapping and comparative genomics using Next-Generation RAD Sequencing of a non-model organism. *PLoS ONE*, **6**, e19315.

Berthelot C, Brunet F, Chalopin D *et al.* (2014) The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. *Nature Communications*, **5**, 3657.

Blischak PD, Kubatko LS, Wolfe AD (2015) Accounting for genotype uncertainty in the estimation of allele frequencies in autopolyploids. *Molecular Ecology Resources*, Early View, doi:10.1111/1755-0998.12493.

Bogart JP, Bi K (2013) Genetic and genomic interactions of animals with different ploidy levels. *Cytogenetic and Genome Research*, **140**, 117–136.

Bottger SA, Eno CC, Walker CW (2011) Methods for generating triploid green sea urchin embryos: an initial step in producing triploid adults for land-based and near-shore aquaculture. *Aquaculture*, **318**, 199–206.

Brieuc MS, Waters CD, Seeb JE, Naish KA (2014) A dense linkage map for Chinook salmon (*Oncorhynchus tshawytscha*) reveals variable chromosomal divergence after an ancestral whole genome duplication event. *G3 (Bethesda)*, **4**, 447–460.

Chapman MA, Abbott RJ (2010) Introgression of fitness genes across a ploidy barrier. *New Phytologist*, **186**, 63–71.

Chapman JA, Mascher M, Buluc AN *et al.* (2015) A whole-genome shotgun approach for assembling and anchoring the hexaploid bread wheat genome. *Genome Biology*, **16**, 26.

Charlesworth B (2009) Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics*, **10**, 195–205.

Chutimanitsakun Y, Nipper RW, Cuesta-Marcos A *et al.* (2011) Construction and application for QTL analysis of a Restriction Site Associated DNA (RAD) linkage map in barley. *BMC Genomics*, **12**, 4.

Clevenger JP, Ozias-Akins P (2015) SWEEP: a tool for filtering high-quality SNPs in polyploid crops. *G3 (Bethesda)*, **5**, 1797–1803.

Comai L (2005) The advantages and disadvantages of being polyploid. *Nature Reviews Genetics*, **6**, 836–846.

Cornejo OE, Fisher D, Escalante AA (2015) Genome-wide patterns of genetic polymorphism and signatures of selection in *Plasmodium vivax*. *Genome Biology and Evolution*, **7**, 106–119.

Dann TH, Habicht C, Baker TT, Seeb JE (2013) Exploiting genetic diversity to balance conservation and harvest of migratory salmon. *Canadian Journal of Fisheries and Aquatic Sciences*, **70**, 785–793.

Danzmann RG, Bogart JP (1982) Evidence for a polymorphism in gametic segregation using a malate dehydrogenase locus in the tetraploid treefrog *Hyla versicolor*. *Genetics*, **100**, 287–306.

Davey JW, Hohenlohe PA, Etter PD, *et al.* (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, **12**, 499–510.

Davidson WS (2013) Understanding salmonid biology from the Atlantic salmon genome. *Genome*, **56**, 548–550.

Davidson WS, Koop BF, Jones SJM *et al.* (2010) Sequencing the genome of the Atlantic salmon (*Salmo salar*). *Genome Biology*, **11**, 403.

Dolezel J, Vrana J, Capal P *et al.* (2014) Advances in plant chromosome genomics. *Biotechnology Advances*, **32**, 122–136.

Dufresne F (2016) Don't throw the baby out with the bathwater: identifying and mapping paralogs in salmonids. *Molecular Ecology Resources*, **16**, 7–9.

Dufresne F, Stift M, Vergilino R, Mable BK (2014) Recent progress and challenges in population genetics of polyploid organisms: an

overview of current state-of-the-art molecular and statistical tools. *Molecular Ecology*, **23**, 40–69.
Eckert AJ, Dyer RJ (2012) Defining the landscape of adaptive genetic diversity. *Molecular Ecology*, **21**, 2836–2838.
Eckert AJ, Bower AD, Gonzalez-Martinez SC *et al.* (2010) Back to nature: ecological genomics of loblolly pine (*Pinus taeda*, Pinaceae). *Molecular Ecology*, **19**, 3789–3805.
Ellegren H (2014) Genome sequencing and population genomics in non-model organisms. *Trends in Ecology & Evolution*, **29**, 51–63.
Ellegren H, Smeds L, Burri R *et al.* (2012) The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature*, **491**, 756–760.
Everett MV, Seeb JE (2014) Detection and mapping of QTL for temperature tolerance and body size in Chinook salmon (*Oncorhynchus tshawytscha*) using genotyping by sequencing. *Evolutionary Applications*, **7**, 480–492.
Fan Y, Shabala S, Ma YL, Xu RG, Zhou MX (2015) Using QTL mapping to investigate the relationships between abiotic stress tolerance (drought and salinity) and agronomic and physiological traits. *BMC Genomics*, **16**, 43.
Fins L, Seeb LW (1986) Genetic-variation in allozymes of western larch. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere*, **16**, 1013–1018.
Fitzpatrick BM, Johnson JR, Kump DK *et al.* (2010) Rapid spread of invasive genes into a threatened native species. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 3606–3610.
Fraser DJ, Weir LK, Bernatchez L, Hansen MM, Taylor EB (2011) Extent and scale of local adaptation in salmonid fishes: review and meta-analysis. *Heredity*, **106**, 404–420.
Gaeta RT, Pires JC (2010) Homoeologous recombination in allopolyploids: the polyploid ratchet. *New Phytologist*, **186**, 18–28.
Gagnaire PA, Normandeau E, Pavey SA, Bernatchez L (2012) Mapping phenotypic, expression and transmission ratio distortion QTL using RAD markers in the Lake whitefish (*Coregonus clupeaformis*). *Molecular Ecology*, **22**, 3036–3048.
Gharbi K, Gautier A, Danzmann RG *et al.* (2006) A linkage map for brown trout (*Salmo trutta*): Chromosome homeologies and comparative genome organization with other salmonid fish. *Genetics*, **172**, 2405–2419.
Gidskehaug L, Kent M, Hayes BJ, Lien S (2011) Genotype calling and mapping of multisite variants using an Atlantic salmon iSelect SNP array. *Bioinformatics*, **27**, 303–310.
Grimholt U, Johansen R, Smith AJ (2009) A review of the need and possible uses for genetically standardized Atlantic salmon (*Salmo salar*) in research. *Laboratory Animals*, **43**, 121–126.
Guyomard R, Boussaha M, Krieg F, Hervet C, Quillet E (2012) A synthetic rainbow trout linkage map provides new insights into the salmonid whole genome duplication and the conservation of synteny among teleosts. *BMC Genetics*, **13**, 15.
Hecht BC, Campbell NR, Holecek DE, Narum SR (2013) Genome-wide association reveals genetic basis for the propensity to migrate in wild populations of rainbow and steelhead trout. *Molecular Ecology*, **22**, 3061–3076.
Hemmer-Hansen J, Nielsen EE, Therkildsen NO *et al.* (2013) A genomic island linked to ecotype divergence in Atlantic cod. *Molecular Ecology*, **22**, 2653–2667.
Hohenlohe PA, Bassham S, Etter PD, *et al.* (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics*, **6**, e1000862.
Hohenlohe PA, Day MD, Amish SJ *et al.* (2013) Genomic patterns of introgression in rainbow and westslope cutthroat trout illuminated by overlapping paired-end RAD sequencing. *Molecular Ecology*, **22**, 3002–3013.
Houston RD, Davey JW, Bishop SC *et al.* (2012) Characterisation of QTL-linked and genome-wide restriction site-associated DNA (RAD) markers in farmed Atlantic salmon. *BMC Genomics*, **13**, 244.
Huang L, Ma F, Chapman A, Lu S, Xie XS (2015) Single-cell whole-genome amplification and sequencing: methodology and applications. *Annual Review of Genomics and Human Genetics*, **16**, 79–102.
Johnston SE, Lindqvist M, Niemela E *et al.* (2013) Fish scales and SNP chips: SNP genotyping and allele frequency estimation in individual and pooled DNA from historical samples of Atlantic salmon (*Salmo salar*). *BMC Genomics*, **14**, 439.
Kodama M, Brieuc MS, Devlin RH, Hard JJ, Naish KA (2014) Comparative mapping between Coho Salmon (*Oncorhynchus kisutch*) and three other salmonids suggests a role for chromosomal rearrangements in the retention of duplicated regions following a whole genome duplication event. *G3 (Bethesda)*, **4**, 1717–1730.
Komen H, Thorgaard GH (2007) Androgenesis, gynogenesis and the production of clones in fishes: a review. *Aquaculture*, **269**, 150–173.
Kondo Y, Kashiwagi A (2004) Experimentally induced autotetraploidy and allotetraploidy in two Japanese pond frogs. *Journal of Herpetology*, **38**, 381–392.
Kooij TWA, Carlton JM, Bidwell SL *et al.* (2005) A Plasmodium whole-genome synteny map: indels and synteny breakpoints as foci for species-specific genes. *Plos Pathogens*, **1**, 349–361.
Krasileva KV, Buffalo V, Bailey P *et al.* (2013) Separating homeologs by phasing in the tetraploid wheat transcriptome. *Genome Biology*, **14**, R66.
Kroeger PT, Poureetezadi SJ, McKee R *et al.* (2014) Production of haploid zebrafish embryos by in vitro fertilization. *Jove-Journal of Visualized Experiments*, **89**, e51708.
Leaché AD, Chavez AS, Jones LN *et al.* (2015) Phylogenomics of phrynosomatid lizards: conflicting signals from sequence capture versus Restriction Site Associated DNA sequencing. *Genome Biology and Evolution*, **7**, 706–719.
Li Q, Hisatsune T, Kijima A (2004) Induction of haploid androgenesis in pacific oyster by UV irradiation. *Marine Biotechnology*, **6**, 291–297.
Lien S, Gidskehaug L, Moen T *et al.* (2011) A dense SNP-based linkage map for Atlantic salmon (*Salmo salar*) reveals extended chromosome homeologies and striking differences in sex-specific recombination patterns. *BMC Genomics*, **12**, 615.
Lighten J, Van Oosterhout C, Paterson IG, McMullan M, Bentzen P (2014) Ultra-deep Illumina sequencing accurately identifies MHC class IIb alleles and provides evidence for copy number variation in the guppy (*Poecilia reticulata*). *Molecular Ecology Resources*, **14**, 753–767.
Limborg MT, Waples RK, Seeb JE, Seeb LW (2014) Temporally isolated lineages of pink salmon reveal unique signatures of selection on distinct pools of standing genetic variation. *Journal of Heredity*, **105**, 741–751.
Limborg MT, Waples RK, Allendorf FW, Seeb JE (2015) Linkage mapping reveals strong chiasma interference in sockeye salmon: implications for interpreting genomic data. *G3 (Bethesda)*, **5**, 2463–2473.
Lindner KR, Seeb JE, Habicht C *et al.* (2000) Gene-centromere mapping of 312 loci in pink salmon by half-tetrad analysis. *Genome*, **43**, 538–549.
Logan-Young CJ, Yu JZ, Verma SK, Percy RG, Pepper AE (2015) SNP discovery in complex allotetraploid genomes (*Gossypium* spp., Malvaceae) using genotyping by sequencing. *Applications in Plant Sciences*, **3**, 1400077.
Mable BK, Alexandrou MA, Taylor MI (2011) Genome duplication in amphibians and fish: an extended synthesis. *Journal of Zoology*, **284**, 151–182.

Macqueen DJ, Johnston IA (2014) A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. *Proceedings of the Royal Society B: Biological Sciences*, **281**, 20132881.

Mascher M, Muehlbauer GJ, Rokhsar DS *et al.* (2013) Anchoring and ordering NGS contig assemblies by population sequencing (POPSEQ). *Plant Journal*, **76**, 718–727.

Mastretta-Yanes A, Zamudio S, Jorgensen TH *et al.* (2014) Gene duplication, population genomics, and species-level differentiation within a tropical mountain shrub. *Genome Biology and Evolution*, **6**, 2611–2624.

May B, Delany ME (2015) Meiotic models to explain classical linkage, pseudolinkage, and chromosomal pairing in tetraploid derivative salmonid genomes: II. Wright is still right. *Journal of Heredity*, **106**, 762–766.

Mayrose I, Zhan SH, Rothfels CJ *et al.* (2011) Recently formed polyploid plants diversify at lower rates. *Science*, **333**, 1257.

McKinney GJ, Seeb LW, Larson WA *et al.* (2015) An integrated linkage map reveals candidate genes underlying adaptive variation in Chinook salmon (*Oncorhynchus tshawytscha*). *Molecular Ecology Resources*, Early View, doi: 10.1111/1755-0998.12479.

Mefford HC, Trask BJ (2002) The complex structure and dynamic evolution of human subtelomeres. *Nature Reviews Genetics*, **3**, 91–102.

Mefford HC, Linardopoulou E, Coil D, van den Engh G, Trask BJ (2001) Comparative sequencing of a multicopy subtelomeric region containing olfactory receptor genes reveals multiple interactions between non-homologous chromosomes. *Human Molecular Genetics*, **10**, 2363–2372.

Molina-Luzon MJ, Hermida M, Navajas-Perez R *et al.* (2015) First haploid genetic map based on microsatellite markers in Senegalese sole (*Solea senegalensis*, Kaup 1858). *Marine Biotechnology*, **17**, 8–22.

Morgan CC, McCartney AM, Donoghue MTA *et al.* (2013) Molecular adaptation of telomere associated genes in mammals. *BMC Evolutionary Biology*, **13**, 251.

Mosher JJ, Bowman B, Bernberg EL *et al.* (2014) Improved performance of the PacBio SMRT technology for 16S rDNA sequencing. *Journal of Microbiological Methods*, **104**, 59–60.

Muhlfeld CC, Kalinowski ST, McMahon TE *et al.* (2009) Hybridization rapidly reduces fitness of a native trout in the wild. *Biology Letters*, rsbl. 2009.0033.

Narum SR, Buerkle CA, Davey JW, Miller MR, Hohenlohe PA (2013) Genotyping-by-sequencing in ecological and conservation genomics. *Molecular Ecology*, **22**, 2841–2847.

Ogugo V, Semagn K, Beyene Y *et al.* (2015) Parental genome contribution in maize DH lines derived from six backcross populations using genotyping by sequencing. *Euphytica*, **202**, 129–139.

Ohno S (1970a) The enormous diversity in genome sizes of fish as a reflection of nature's extensive experiments with gene duplication. *Transactions of the American Fisheries Society*, **99**, 120–130.

Ohno S (1970b) *Evolution by Gene Duplication*. George Alien & Unwin Ltd, London. Springer-Verlag, Berlin, Heidelberg and New York.

Otto SP, Whitton J (2000) Polyploid incidence and evolution. *Annual Review of Genetics*, **34**, 401–437.

Palti Y, Gao G, Liu S *et al.* (2015) The development and characterization of a 57K single nucleotide polymorphism array for rainbow trout. *Molecular Ecology Resources*, **15**, 662–672.

Parisod C, Holderegger R, Brochmann C (2010) Evolutionary consequences of autopolyploidy. *New Phytologist*, **186**, 5–17.

Parsons JE, Thorgaard GH (1985) Production of androgenetic diploid rainbow-trout. *Journal of Heredity*, **76**, 177–181.

Poland JA, Rife TW (2012) Genotyping-by-sequencing for plant breeding and genetics. *Plant Genome*, **5**, 92–102.

Poland JA, Brown PJ, Sorrells ME, Jannink JL (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS ONE*, **7**, e32253.

Ramsey J, Schemske DW (1998) Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annual Review of Ecology and Systematics*, **29**, 467–501.

Richier S, Kerros ME, de Vargas C *et al.* (2009) Light-dependent transcriptional regulation of genes of biogeochemical interest in the diploid and haploid life cycle stages of *Emiliania huxleyi*. *Applied and Environmental Microbiology*, **75**, 3366–3369.

Robertson LS, Cornman RS (2014) Transcriptome resources for the frogs *Lithobates clamitans* and *Pseudacris regilla*, emphasizing antimicrobial peptides and conserved loci for phylogenetics. *Molecular Ecology Resources*, **14**, 178–183.

Rondeau EB, Minkley DR, Leong JS *et al.* (2014) The genome and linkage map of the Northern pike (*Esox lucius*): conserved synteny revealed between the salmonid sister group and the Neoteleostei. *PLoS ONE*, **9**, e102089.

Sakamoto T, Danzmann RG, Gharbi K *et al.* (2000) A microsatellite linkage map of rainbow trout (*Oncorhynchus mykiss*) characterized by large sex-specific differences in recombination rates. *Genetics*, **155**, 1331–1345.

Santure AW, De Cauwer I, Robinson MR *et al.* (2013) Genomic dissection of variation in clutch size and egg mass in a wild great tit (*Parus major*) population. *Molecular Ecology*, **22**, 3949–3962.

Savage AE, Kiemnec-Tyburczy KM, Ellison AR, Fleischer RC, Zamudio KR (2014) Conservation and divergence in the frog immunome: pyrosequencing and de novo assembly of immune tissue transcriptomes. *Gene*, **542**, 98–108.

Scheerer PD, Thorgaard GH, Allendorf FW (1991) Genetic analysis of androgenetic rainbow trout. *The Journal of Experimental Zoology*, **260**, 382–390.

Seeb JE, Seeb LW (1986) Gene mapping of isozyme loci in chum salmon. *Journal of Heredity*, **77**, 399–402.

Seeb JE, Pascal CE, Grau ED *et al.* (2011) Transcriptome sequencing and high-resolution melt analysis advance single nucleotide polymorphism discovery in duplicated salmonids. *Molecular Ecology Resources*, **11**, 335–348.

Selman GG (1958) An ultra-violet light method for producing haploid amphibian embryos. *Journal of Embryology and Experimental Morphology*, **6**, 634–637.

Selmecki AM, Maruvka YE, Richmond PA *et al.* (2015) Polyploidy can drive rapid adaptation in yeast. *Nature*, **519**, 349–352.

Shaffer HB, Gidiş M, McCartney-Melstad E *et al.* (2015) Conservation genetics and genomics of amphibians and reptiles. *Annual Review of Animal Biosciences*, **3**, 113–138.

Siger A, Michalak M, Cegielska-Taras T *et al.* (2015) Genotype and environment effects on tocopherol and plastochromanol-8 contents of winter oilseed rape doubled haploid lines derived from F1 plants of the cross between yellow and black seeds. *Industrial Crops and Products*, **65**, 134–141.

Sork VL, Aitken SN, Dyer RJ *et al.* (2013) Putting the landscape into the genomics of trees: approaches for understanding local adaptation and population responses to changing climate. *Tree Genetics & Genomes*, **9**, 901–911.

Spruell P, Pilgrim KL, Greene BA *et al.* (1999) Inheritance of nuclear DNA markers in gynogenetic haploid pink salmon. *Journal of Heredity*, **90**, 289–296.

Stift M, Berenos C, Kuperus P, van Tienderen PH (2008) Segregation models for disomic, tetrasomic and intermediate inheritance

in tetraploids: a general procedure applied to Rorippa (Yellow cress) microsatellite data. *Genetics*, **179**, 2113–2123.

Stölting KN, Nipper R, Lindtke D *et al.* (2013) Genomic scan for single nucleotide polymorphisms reveals patterns of divergence and gene flow between ecologically divergent species. *Molecular Ecology*, **22**, 842–855.

Thorgaard GH (1992) Application of genetic technologies to rainbow trout. *Aquaculture*, **100**, 85–97.

Thorgaard GH, Allendorf FW, Knudsen KL (1983) Gene-centromere mapping in rainbow trout: high interference over long map distances. *Genetics*, **103**, 771–783.

Thorgaard G, Bailey G, Williams D *et al.* (2002) Status and opportunities for genomics research with rainbow trout. *Comp Biochem Physiol B Biochem Mol Biol*, **133**, 609–646.

Tomiuk J, Guldbrandtsen B, Loeschcke V (2009) Genetic similarity of polyploids: a new version of the computer program POPDIST (version 1.2.0) considers intraspecific genetic differentiation. *Molecular Ecology Resources*, **9**, 1364–1368.

Voorrips RE, Gort G, Vosman B (2011) Genotype calling in tetraploid species from bi-allelic marker data using mixture models. *BMC Bioinformatics*, **12**, 172.

Wallace JG, Larsson SJ, Buckler ES (2014) Entering the second century of maize quantitative genetics. *Heredity*, **112**, 30–38.

Wang J, Scribner KT (2014) Parentage and sibship inference from markers in polyploids. *Molecular Ecology Resources*, **14**, 541–553.

Waples RK, Seeb LW, Seeb JE (2016) Linkage mapping with paralogs exposes regions of residual tetrasomic inheritance in chum salmon (*Oncorhynchus keta*). *Molecular Ecology Resources*, **16**, 17–28.

Warren IA, Ciborowski KL, Casadei E *et al.* (2014) Extensive local gene duplication and functional divergence among paralogs in Atlantic salmon. *Genome Biology and Evolution*, **6**, 1790–1805.

Wolfe KH (2001) Yesterday's polyploids and the mystery of diploidization. *Nature Reviews Genetics*, **2**, 333–341.

Wright JE, Johnson K, Hollister A, May B (1983) Meiotic models to explain classical linkage, pseudolinkage, and chromosome pairing in tetraploid derivative salmonid genomes. In: *Isozymes: Current Topics in Biological and Medical Research* (eds Rattazzi MC, Scandalios JG, Whitt GS), pp. 239–260. Alan R. Liss, New York.

Wu G, Hufnagel DE, Denton AK, Shiu S-H (2015) Retained duplicate genes in green alga *Chlamydomonas reinhardtii* tend to be stress responsive and experience frequent response gains. *BMC Genomics*, **16**, 149.

Xu P, Zhang X, Wang X *et al.* (2014) Genome sequence and genetic diversity of the common carp, *Cyprinus carpio*. *Nature Genetics*, **46**, 1212–1219.

---

M.T.L., L.W.S. and J.E.S. conceived the study and wrote the manuscript. All authors approved the final version.

---

## Data accessibility

The family data underlying Fig. 1b are available from the original study by Waples *et al.* (2016) at the DRYAD data repository: doi:10.5061/dryad.5b64r.2.

doi: 10.1111/mec.13601